October 24, 2005
D. Jacob Wildstrom

Convergence on Directed Graphs

From Perron-Frobenius to Convergence

1 Introduction

Working with a graph G and a probability transition matrix P, we previously saw that the criterion for convergence was that G is strongly connected and the the GCD of the cycle lengths in G is 1. We shall prove that this condition is indeed sufficient to demonstrate convergence. By the Perron-Frobenius Theorem, this condition is equvalent to the statement that the norms of all eigenvalues of P except for $\rho_0 = 1$ are less than 1. We shall thus show, from a spectral argument alone, that the random walk converges.

From an algebraic perspective, what we want to do is characterize P^t. Ideally we want to characterize this in terms of eigenvalues alone; however, some matrices are not amenable to such characterization. For instance, $P = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not diagonalizable, so P^t would not be describable in terms of the eignevalues alone. However, this matrix also does not correspond to a strongly connected digraph: nonetheless, we cannot assume diagonalizability when studying the transition matrices of strongly connected graphs.

Exercise: Find a strongly connected G such that the associated probability transition matrix P does not have n distinct eigenvectors; that is to say, that P is undiagonalizable.

In general, we can investigate convergence over a class of similar matrices: M and SMS^{-1} have the same long-term convergence profile, since $(SMS^{-1})^t = SM^tS^{-1}$. In addition, characterizing in terms of eigenvalues is supported by the similarity relation, since M and SMS^{-1} have the same eigenvalues. It is worth noting, however, that even though P represents transitions on a strongly connected graph, SPS^{-1} may not.

2 Jordan matrix decomposition

Were P diagonalizable, we would simply note the similarity of P to a matrix with ρ_i on the diagonal, and since $\rho_i^t \to 0$ for all $i \neq 0$, P^t would necessarily converge. This approach works for undirected graphs: even though P itself would not necessarily be symmetric, it would be similar to the symmetric matrix $D^{\frac{1}{2}}PD^{-\frac{1}{2}}$, which is diagonalizable, as are all symmetric matrices.

However, with a directed graph G, we are not guaranteed that the probability transition matrix P is diagonalizable, and must thus use a weaker form than diagonal form. This is a purpose for which *Jordan matrix decomposition* is ideal:

Theorem 1. *Any matrix M can be decomposed into the form $S\Gamma S^{-1}$, where S is an invertable matrix, and Γ has the block form:*

$$\Gamma = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_k \end{bmatrix}$$

with sub-blocks of the form

$$B_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix}$$

where $\lambda_1, \ldots, \lambda_k$ are the eigenvalues of M.

Let us start by taking $P_0 = \vec{1}\phi$, where ϕ is the row eigenvector of P associated with the eigenvalue $\rho_0 = 1$ with positive entries which sum to 1, called the *Perron vector*, and let $M = P - P_0$. This matrix has the same spectral character as P except that the eigenvalue ρ_0 is replaced with 0, so we may show that M^t in fact converges to the zero matrix. We than can demonstrate that $(P - P_0)^t = P^t - P_0$, so that M's convergence to zero implies P's convergence to P_0.

Exercise: Show that $(P - P_0)^t = P^t - P_0$.

Decomposing M this way, we find that $M^t = S\Gamma^t S^{-1}$. Since Γ is block-diagonal,

$$\Gamma^t = \begin{bmatrix} B_1^t & 0 & \cdots & 0 \\ 0 & B_2^t & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_k^t \end{bmatrix},$$

so determining Γ^t depends on the behavior of each B_i^t. Let us note that $B_i = \lambda_i I + F$, where F is the matrix with zero-entries everywhere except in the cells directly above the main diagonal (i.e., the super-diagonal), which contain 1. Then, we use the binomial expansion of this sum. In general doing so could be quite ugly since matrices do not commute, but since one of the terms in the sum is $\lambda_i I$, which does commute, our binomial expansion is the traditional commutative binomial formula:

$$B_i^t = (\lambda_i I + F)^t = \lambda_i^t I + \binom{t}{1}\lambda_i^{t-1}F + \binom{t}{2}\lambda_i^{t-2}F^2 + \cdots$$

For F an $m \times m$ matrix, $F^m = 0$, so this sum can be rewritten as

$$\lambda_i^{t-m+1}\left(\lambda_i^{m-1}I + \binom{t}{1}\lambda_i^{m-2}F + \binom{t}{2}\lambda_i^{m-3}F^2 + \cdots + \binom{t}{m-1}\lambda_i^0 F^{m-1}\right)$$

Note that the expression in the parentheses is a matrix in which each entry is an $(m-1)$th-degree polynomial in t; since $|\lambda_i| < 1$, the product is dominated by the λ_i^{t-m+1} term,which approaches zero as $t \to \infty$. Thus $B_i^t \to 0$, so by the block structure $\Gamma^t \to 0$ and thus $M^t \to 0$.

3 Four notions of convergence

However, we care not simply about convergence, but convergence rates, and not simply about convergence of the matrix P^t, but of the probability distribution fP^t. There are four different metrics under which convergence is measured, arranged here from least easily demonstrated to most demonstrable:

Convergence in the $\mathcal{L}^2$ norm This is the familiar Euclidean metric for convergence,

$$\Delta_{\mathcal{L}^2}(t) = \max_f \|fP^t - \phi\| = \max_f \sqrt{\sum_x [(fP^t)(x) - \phi(x)]^2},$$

which is well-known, but for our purposes too weak, inasmuch as it doesn't capture the convergence of the distribution at every vertex.

Total-variation convergence The *total-variation norm* is a metric favored by Persi Diaconis [1]:

$$\Delta_{tv}(t) = \max_{A \subset V(G)} \max_x \left| \sum_{y \in A} P^t(x,y) - \phi(y) \right|$$

Exercise: Show that $\Delta_{tv}(t) = \frac{1}{2} \max_x \sum_y |P^t(x,y) - \phi(y)|$.

χ-square convergence This metric utilizes variation in each component of a vector, in a manner similar to the $\mathcal{L}^2$-norm, but normalized in each component:

$$\Delta'(t) = \max_x \sqrt{\sum_y \frac{(P^t(x,y) - \phi(y))^2}{\phi(y)}}$$

Relative pointwise distance The simplest of the metrics computationally, relative pointwise distance simply determines the furthest relative distance between two vectors in a single component:

$$\Delta(t) = \max_{x,y} \frac{|P^t(x,y) - \phi(y)|}{\phi(y)}$$

In all of these metrics, the convergence rate is proportional to $\frac{1}{\lambda}$, where λ is the largest eigenvalue of M. We shall demonstrate the convergence explicitly in the total-variation and χ-square distance.

$$\begin{aligned}
\Delta_{tv}(t) &= \frac{1}{2} \max_x \sum_y |P^t(x,y) - \phi(y) \\
&\le \frac{1}{2} \max_x \left(\sum_y \frac{|P^t(x,y) - \phi(y)|^2}{\phi(y)} \right)^{\frac{1}{2}} \quad \text{by the Cauchy-Schwarz Inequality} \\
&\le \frac{1}{2} \Delta'(t)
\end{aligned}$$

We have shown that the total-variation distance is bounded above by half the χ-square distance, so next we shall demonstrate χ-square convergence. Below we use χ_x as the "incidator vector" on element x; that is, $\chi_x(y)$ is 1 for $x = y$ and zero elsewhere. We shall also let Φ be the matrix with the elements of ϕ along its diagonal, and this yields a new representation of the χ-square difference:

$$\begin{aligned}
\Delta'(t) &= \frac{1}{2}\max_x \|\chi_x(P - P_0)^t \Phi^{-\frac{1}{2}}\| \\
&= \frac{1}{2}\max_x \|\chi_x S \Gamma^t S^{-1} \Phi^{-\frac{1}{2}}\| \\
&\leq \frac{1}{2}\|S\| \cdot \|S^{-1}\Phi^{-\frac{1}{2}}\| \cdot \|\Gamma^t\| \\
&\leq \frac{1}{2}\|S\| \cdot \|S^{-1}\Phi^{-\frac{1}{2}}\| \cdot n t^n \rho_1^{t-n} \\
&\leq C t^n \rho_1^{t-n}
\end{aligned}$$

and since $\rho_1 < 1$, this is dominated by ρ_1^{t-n}, which approaches zero.

References

[1] P. Diaconis, *Group Representations in Probability and Statistics*, IMS Lecture Series volume 11, Institute of Mathematical Statistics, Hayward, California, 1988.